



Reconocimiento de Patrones

Version 2022-2

PCA Principal Component Analysis

[Capítulo 3]

Dr. José Ramón Iglesias

DSP-ASIC BUILDER GROUP

Director Semillero TRIAC

Ingeniería Electronica

Universidad Popular del Cesar

Repaso

Valores Propios: λ_l

\mathbf{C} : Matriz de $L \times L$

$$|\mathbf{C} - \lambda_l \mathbf{I}| = 0 \quad \lambda_l : \text{valores propios de } \mathbf{C}$$
$$l = 1, 2, \dots, L$$

$$\{\lambda_l\} = e_{val}(\mathbf{C})$$

Ejemplo:

calcular valores propios de $\mathbf{C} = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}$

$$|\mathbf{C} - \lambda_l \mathbf{I}| = 0$$

$$|\mathbf{C} - \lambda \mathbf{I}| = \begin{vmatrix} 4 - \lambda & 3 \\ 1 & 2 - \lambda \end{vmatrix} = (4 - \lambda)(2 - \lambda) - 3$$

$$\lambda^2 - 6\lambda + 5 = 0 \quad \lambda_1 = 5, \lambda_2 = 1.$$

Vectores Propios: \mathbf{v}_l

$$\mathbf{C}\mathbf{v}_l = \lambda_l \mathbf{v}_l \quad \mathbf{v}_l: \text{vectores propios de } L \times 1$$

Como hay infinitas soluciones: s.t. $\|\mathbf{v}_l\| = 1$

$\{\mathbf{v}_l\}$ son vectores ortonormales:

$$\mathbf{v}_l^\top \mathbf{v}_k = \begin{cases} 1 & \text{si } l = k \\ 0 & \text{en caso contrario} \end{cases}$$

$$\{\mathbf{v}_l\} = e_{vec}(\mathbf{C})$$

Matriz de Vectores Propios: \mathbf{A}

$$\mathbf{A} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_L] : \text{Matriz de } L \times L$$

$$\mathbf{A} = e_{mat}(\mathbf{C})$$

Propiedades:

$$\mathbf{A}\mathbf{A}^T = \mathbf{I} \quad \longleftarrow \quad \mathbf{v}_l^T \mathbf{v}_k = \begin{cases} 1 & \text{si } l = k \\ 0 & \text{en caso contrario} \end{cases}$$

$$\mathbf{C}\mathbf{A} = [\lambda_1 \mathbf{v}_1 \ \lambda_2 \mathbf{v}_2 \ \cdots \ \lambda_L \mathbf{v}_L] \quad \longleftarrow \quad \mathbf{C}\mathbf{v}_l = \lambda_l \mathbf{v}_l$$

Transformaciones Lineales

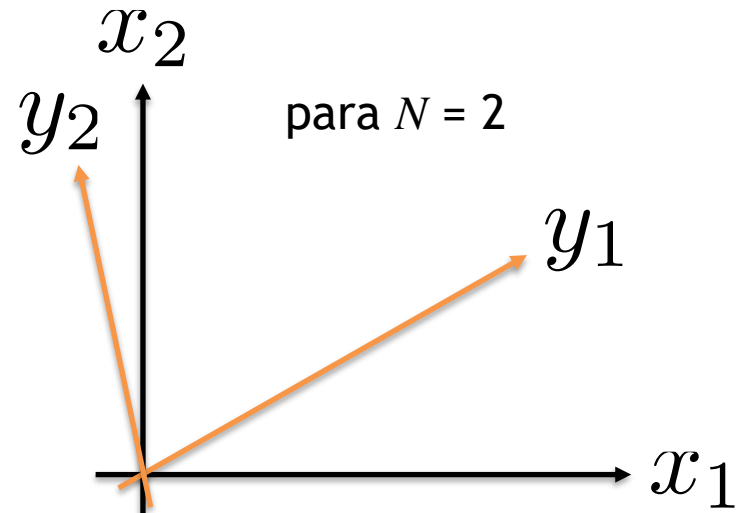
$$y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1N}x_N$$

$$y_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2N}x_N$$

:

$$y_N = a_{N1}x_1 + a_{N2}x_2 + \cdots + a_{NN}x_N$$

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$



- Ángulo entre y_1 e y_2 no necesariamente es 90°
- En esta transformación el origen coincide
- Si se necesita desplazar origen es necesario agregar $+a_{i0}$ a la fila i

Promedio y Varianza:

$$\mathbf{x} = [x_1 \ x_2 \ \cdots \ x_N]^T \quad \text{de } N \times 1$$

(promedio) $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$

(varianza) $\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^N (\bar{x} - x_i)^2$

Covarianza:

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_L] \quad \text{de } N \times L$$

$$\mathbf{M}_x = \underbrace{\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}}_{N \text{ unos}} \underbrace{\begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_L \end{bmatrix}}_{\bar{\mathbf{X}} \text{ promedio de cada columna de } \mathbf{X}}$$

Matriz de $N \times L$: cada columna tiene el promedio de la columna correspondiente de \mathbf{X} .

Covarianza:

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_L] \quad \text{de } N \times L$$

$$\mathbf{M}_x = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} [\bar{x}_1 \ \bar{x}_2 \ \cdots \ \bar{x}_L]$$

$$\mathbf{X}_0 = \mathbf{X} - \mathbf{M}_x \quad \text{cada columna de } \mathbf{X}_0 \text{ tiene media cero}$$

$$\mathbf{C}_x = \frac{1}{N-1} \mathbf{X}_0^\top \mathbf{X}_0 = \text{cov}(\mathbf{X})$$

Covarianza:

$$C_x(l, l) = \sigma_{x_l}^2$$

$$C_x(l, k) = \text{Covarianza de } \mathbf{X}_l \text{ con } \mathbf{X}_k = \text{cov}(\mathbf{x}_k, \mathbf{x}_l)$$

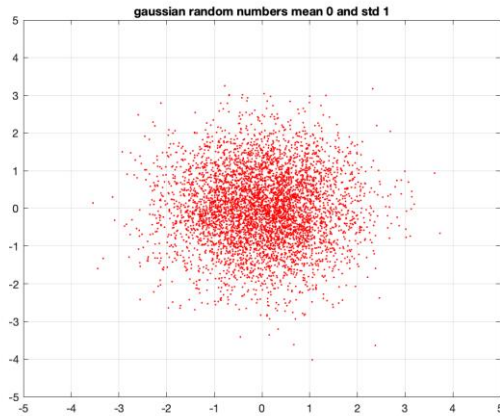
> 0 correlación positiva

$= 0$ no hay correlación

< 0 correlación negativa

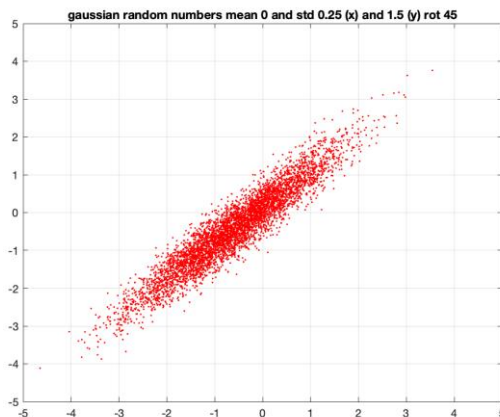
$$\mathbf{C}_x = \frac{1}{N-1} \mathbf{X}_0^T \mathbf{X}_0 = \text{cov}(\mathbf{X})$$

Ejemplo: Matrices de Covarianza



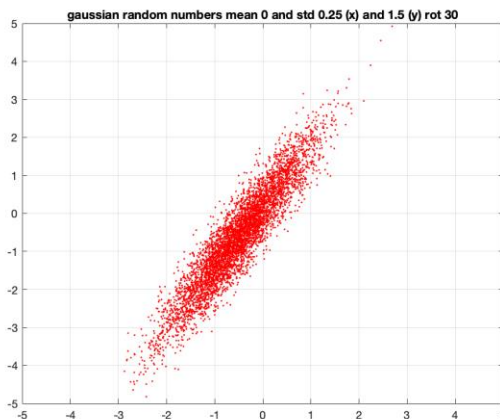
C =

0.9735	0.0041
0.0041	1.0105



C =

1.1465	1.0891
1.0891	1.1604

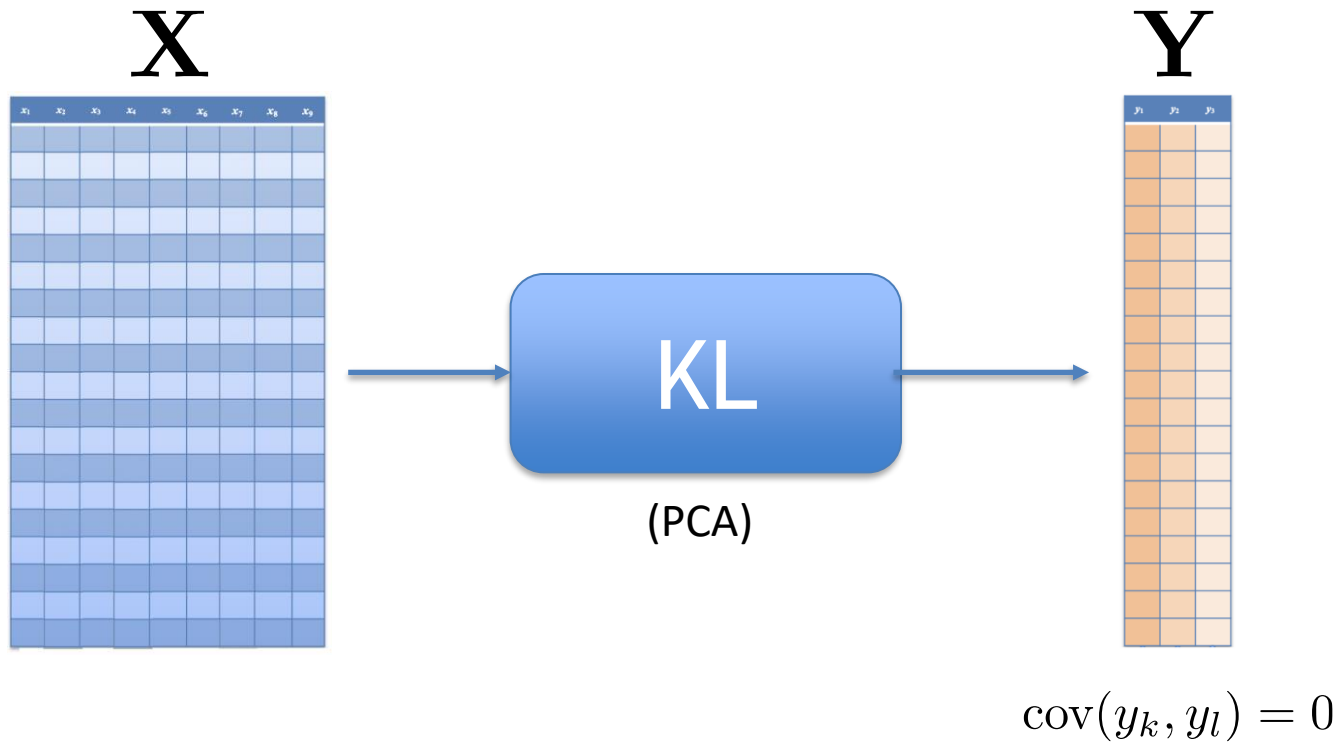


C =

0.6013	0.9290
0.9290	1.6678

TRANSFORMACIÓN
KARHUNEN-LOEVE (PCA)
(1947)

La Transformada de Karhunen-Loeve (KL) es conocida como PCA (Principal Component Analysis). La idea es transformar los datos originales \mathbf{X} en datos nuevos \mathbf{Y} que no tengan correlación entre ellos.



Transformación KL:

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_L] \quad \text{de } N \times L$$

$$\mathbf{X}_0 = \mathbf{X} - \mathbf{M}_x$$

$$\mathbf{C}_x = \frac{1}{N-1} \mathbf{X}_0^\top \mathbf{X}_0 = \text{cov}(\mathbf{X})$$

$$\mathbf{A} = e_{mat}(\mathbf{C}_x)$$

$$\mathbf{Y} = \mathbf{X}_0 \mathbf{A} \quad \leftarrow \text{Transformación KL de } \mathbf{X}$$

$$\mathbf{X} = \mathbf{Y} \mathbf{A}^\top + \mathbf{M}_x \quad \leftarrow \text{Transformación KL inversa de } \mathbf{Y}$$

(\mathbf{A} es ortonormal)

Interpretación de KL:

Covarianza de \mathbf{Y} :

$$\mathbf{C}_y = \frac{1}{N-1} \mathbf{Y}_0^\top \mathbf{Y}_0 = \text{cov}(\mathbf{Y})$$

$$\mathbf{Y}_0 = \mathbf{Y} - \cancel{\mathbf{M}_y} \quad \text{Como } \mathbf{Y} = \mathbf{X}_0 \mathbf{A} \text{ y la media de las} \\ \text{columnas de } \mathbf{X}_0 \text{ es cero, entonces } \mathbf{M}_y = \mathbf{0}$$

$$\mathbf{Y}_0 = \mathbf{Y}$$

$$\mathbf{C}_y = \frac{1}{N-1} \mathbf{Y}^\top \mathbf{Y} = \text{cov}(\mathbf{Y})$$

Interpretación de KL:

Covarianza de \mathbf{Y} :

$$= \frac{1}{N-1} (\mathbf{X}_0 \mathbf{A})^\top \mathbf{X}_0 \mathbf{A}$$

$$= \frac{1}{N-1} \mathbf{A}^\top \mathbf{X}_0^\top \mathbf{X}_0 \mathbf{A}$$

$$\mathbf{C}_y \equiv \frac{\mathbf{A}^\top \mathbf{C}_x \mathbf{A}}{N-1} = \text{cov}(\mathbf{Y})$$

Interpretación de KL:

Covarianza de \mathbf{Y} :

como $\mathbf{A} = e_{mat}(\mathbf{C}_x)$ entonces

$$\mathbf{A} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_L]$$

$$\mathbf{C}_x \mathbf{A} = [\lambda_1 \mathbf{v}_1 \ \lambda_2 \mathbf{v}_2 \ \cdots \ \lambda_L \mathbf{v}_L]$$

$$\mathbf{C}_y \mathbf{G}_y = [\mathbf{v}_1 \mathbf{A} \mathbf{v}_2 \mathbf{C}_x \mathbf{A} \mathbf{v}_L]^\top [\lambda_1 \mathbf{v}_1 \ \lambda_2 \mathbf{v}_2 \ \cdots \ \lambda_L \mathbf{v}_L]$$

Interpretación de KL:

Covarianza de \mathbf{Y} :

$$\mathbf{C}_y = \begin{bmatrix} \lambda_1 \mathbf{v}_1^T \mathbf{v}_1 & \lambda_2 \mathbf{v}_1^T \mathbf{v}_2 & \dots & \lambda_L \mathbf{v}_1^T \mathbf{v}_L \\ \lambda_1 \mathbf{v}_2^T \mathbf{v}_1 & \lambda_2 \mathbf{v}_2^T \mathbf{v}_2 & \dots & \lambda_L \mathbf{v}_2^T \mathbf{v}_L \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 \mathbf{v}_L^T \mathbf{v}_1 & \lambda_2 \mathbf{v}_L^T \mathbf{v}_2 & \dots & \lambda_L \mathbf{v}_L^T \mathbf{v}_L \end{bmatrix}$$
$$\mathbf{C}_y = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_L]^T [\lambda_1 \mathbf{v}_1 \ \lambda_2 \mathbf{v}_2 \ \cdots \ \lambda_L \mathbf{v}_L]$$

Interpretación de KL:

Covarianza de \mathbf{Y} :

$$\mathbf{C}_y = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_L]^\top [\lambda_1 \mathbf{v}_1 \ \lambda_2 \mathbf{v}_2 \ \cdots \ \lambda_L \mathbf{v}_L]$$

$$\mathbf{C}_y = \begin{bmatrix} \lambda_1 \mathbf{v}_1^\top \mathbf{v}_1 = 1 & 0 & \cdots & 0 \\ 0 & \lambda_2 \mathbf{v}_2^\top \mathbf{v}_2 = 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_L \mathbf{v}_L^\top \mathbf{v}_L = 1 \end{bmatrix}$$

Interpretación de KL:

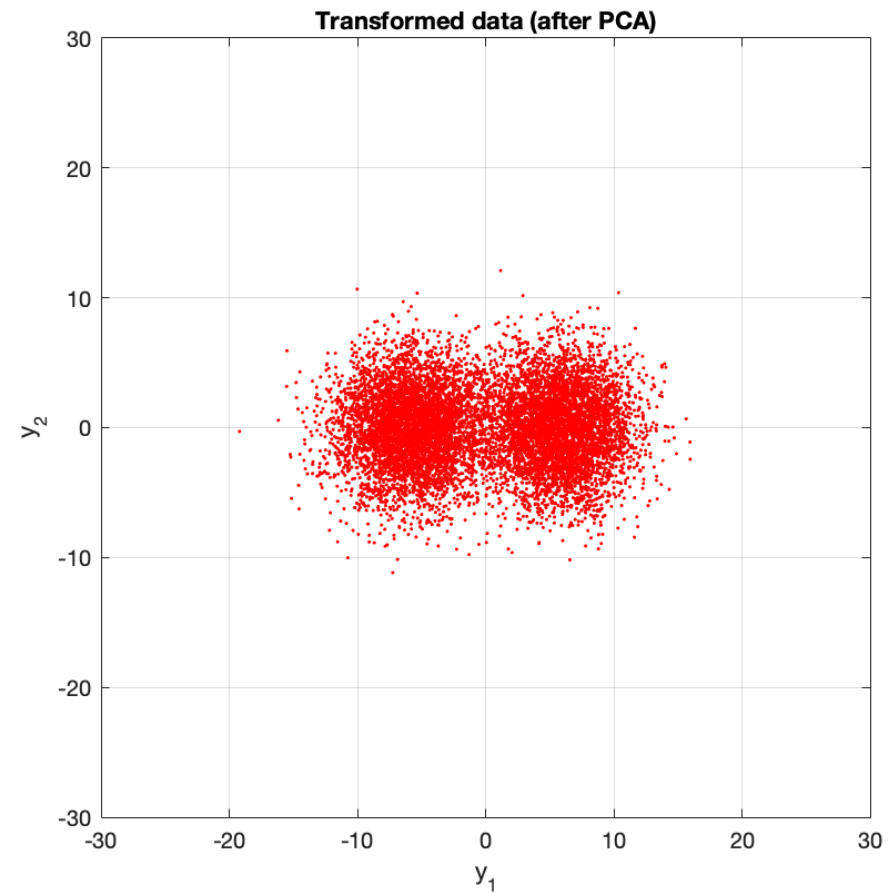
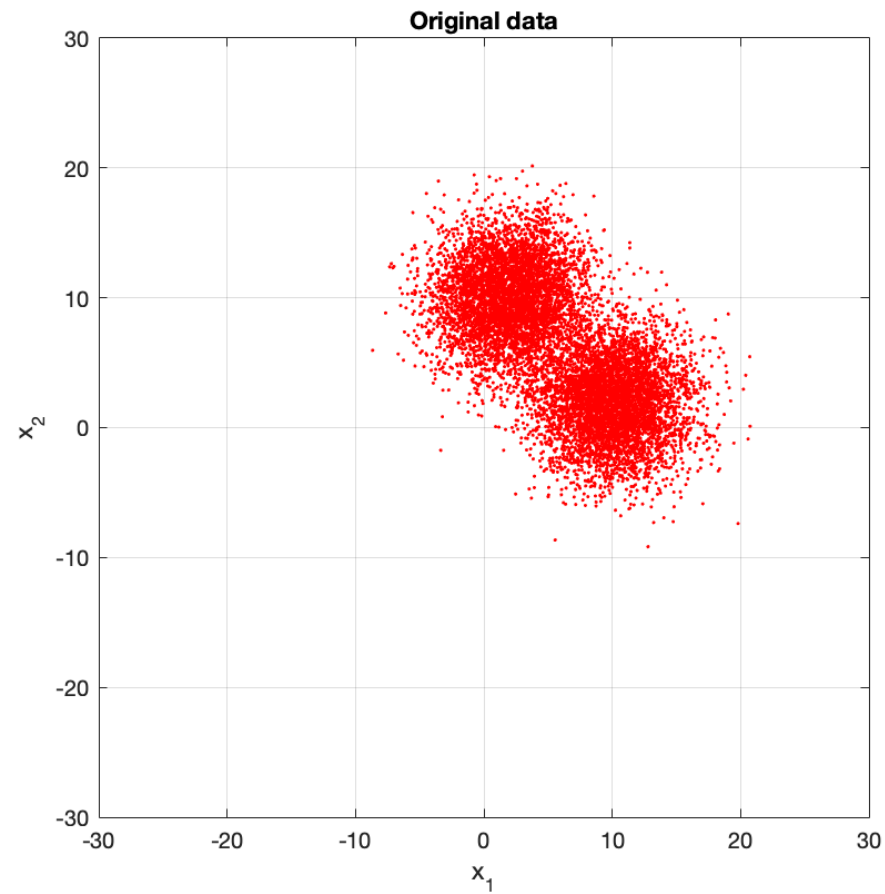
Covarianza de \mathbf{Y} :

$$\mathbf{C}_y = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & \vdots & & \\ 0 & 0 & \dots & \lambda_L \end{bmatrix}$$

Generalmente, se ordena de tal forma que: $\lambda_1 > \lambda_2 > \dots > \lambda_L$

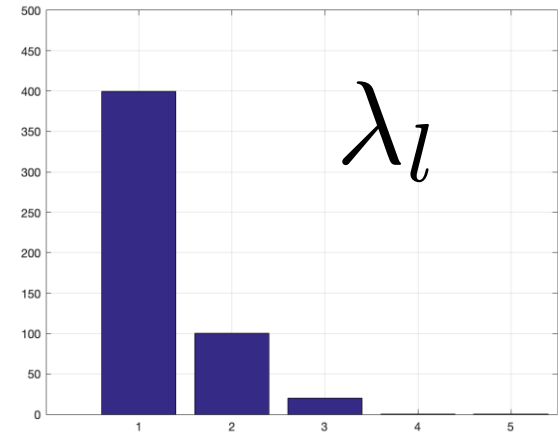
SIGNIFICADO

Ejemplo en 2D

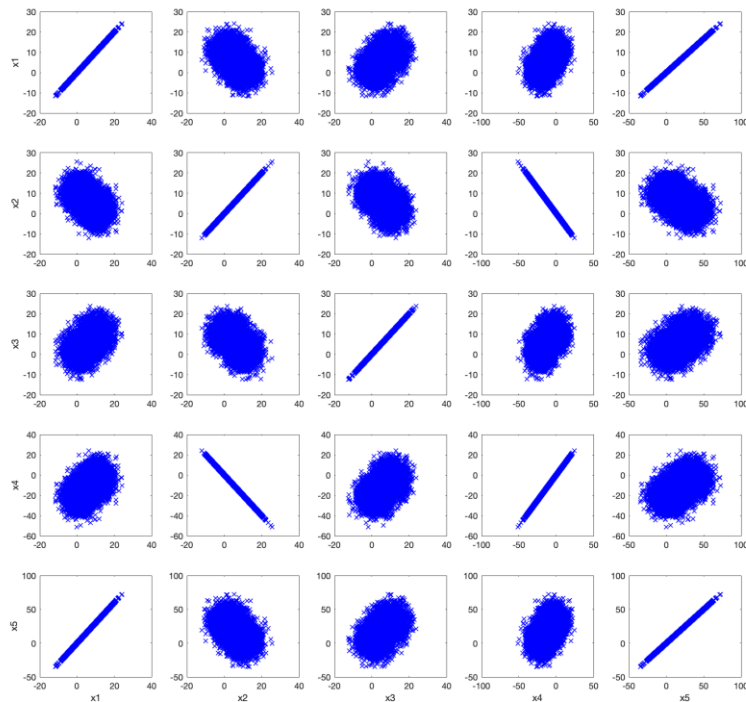


X

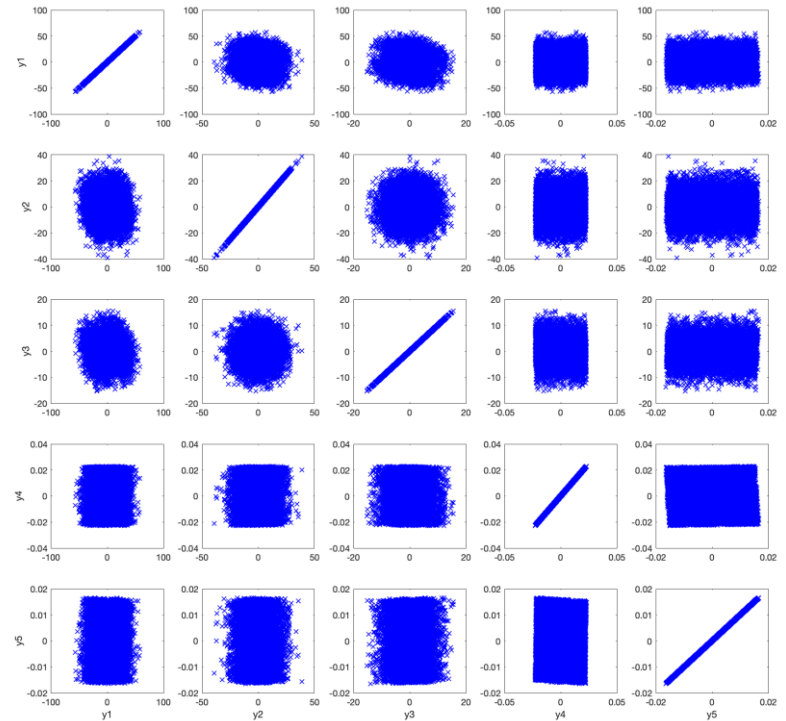
Ejemplo en 5D



X

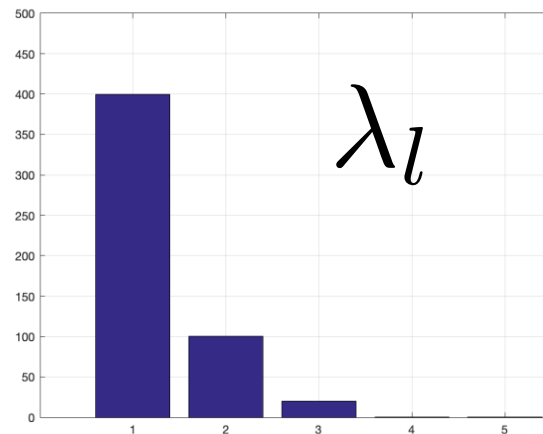


Y

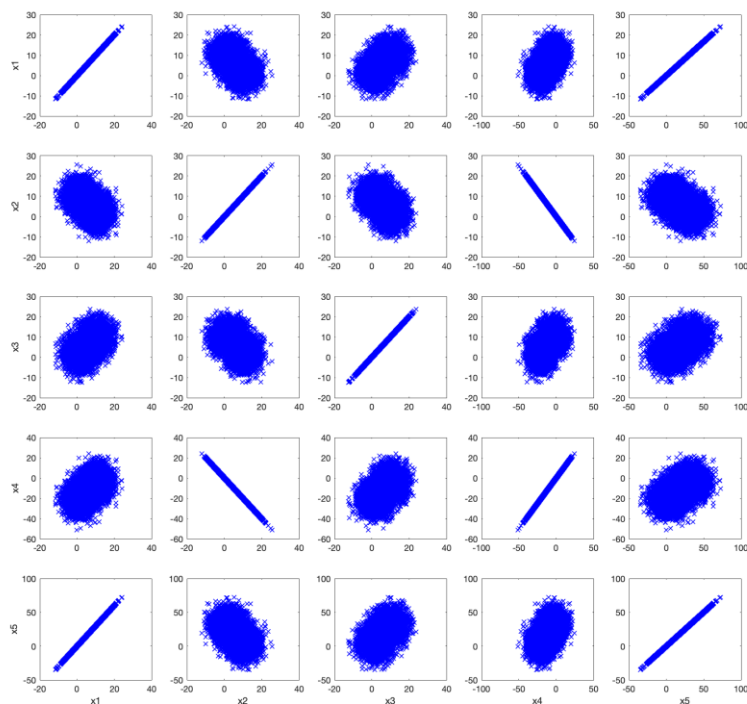


Y

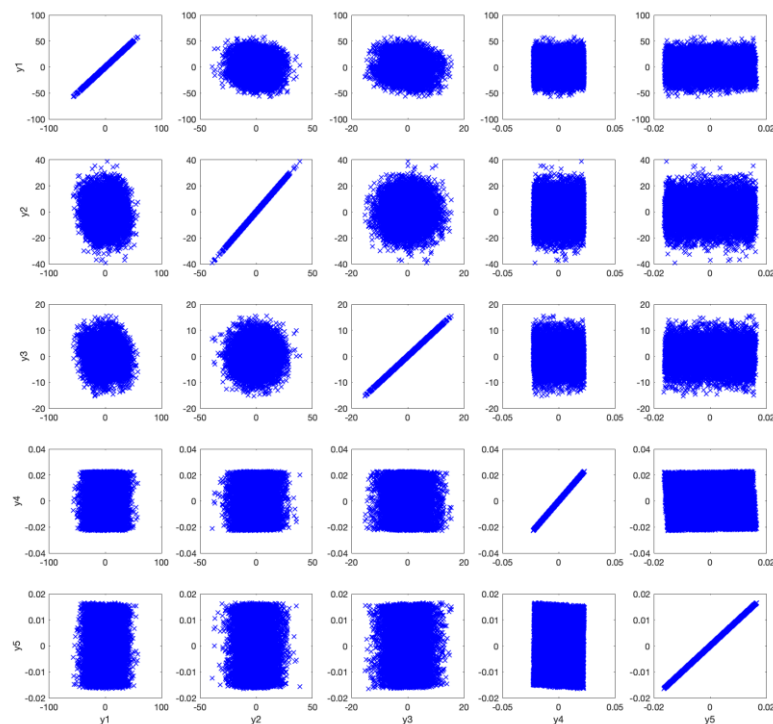
Ejemplo en 5D



X

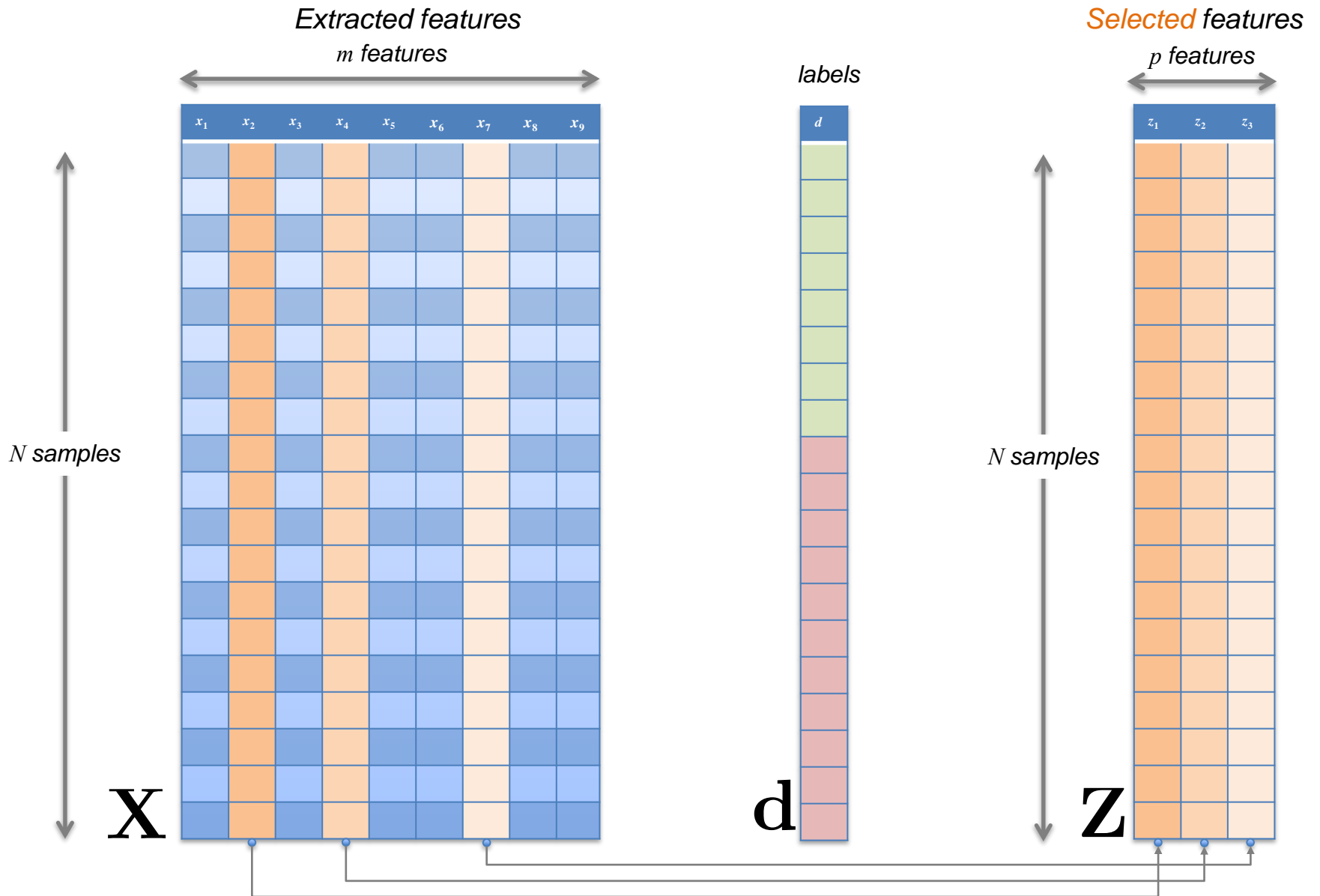


Y

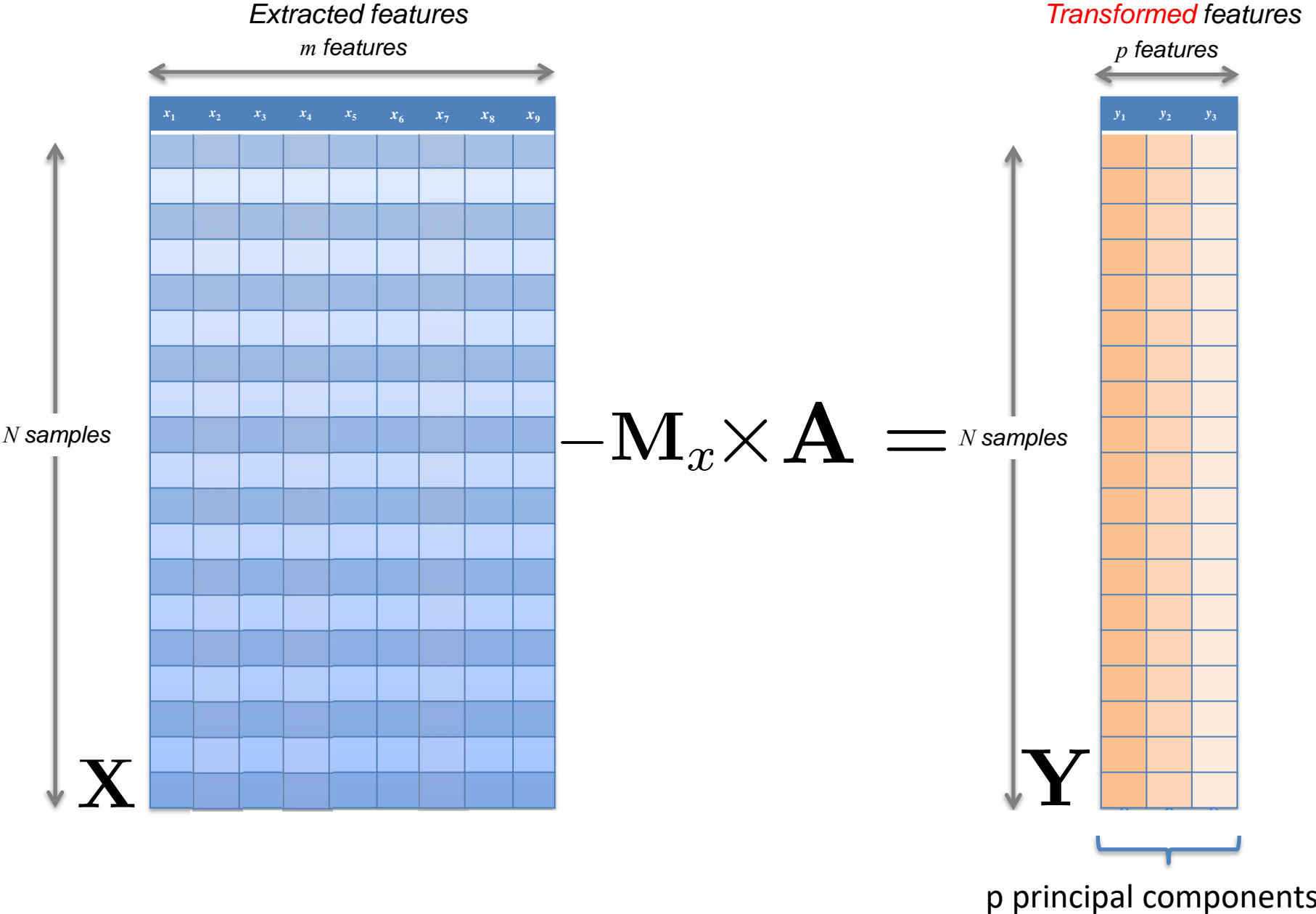


SIGNIFICADO EN LA TRANSFORMACIÓN DE CARACTERÍSTICAS

SELECCION DE CARACTERISTICAS



PCA

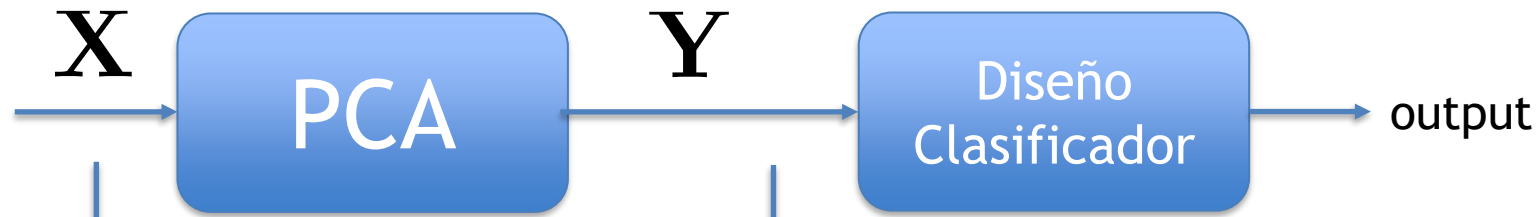


Significado de PCA:

1. **\mathbf{Y}** es una nueva representación de **\mathbf{X}**
(es una transformación lineal de **\mathbf{X}**)
2. Las columnas de **\mathbf{Y}** no tienen correlación: $C_y(k, l) = 0$ para $k \neq l$
3. Ordenado de mayor a menor las columnas de **\mathbf{Y}** están ordenadas por importancia

APLICACIÓN A RECONOCIMIENTO DE PATRONES

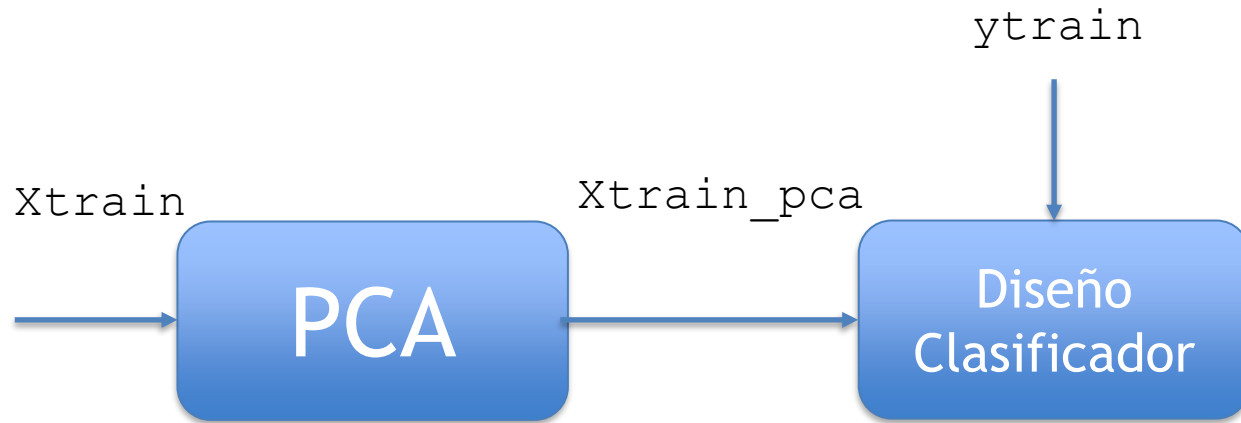
Estrategia con PCA:



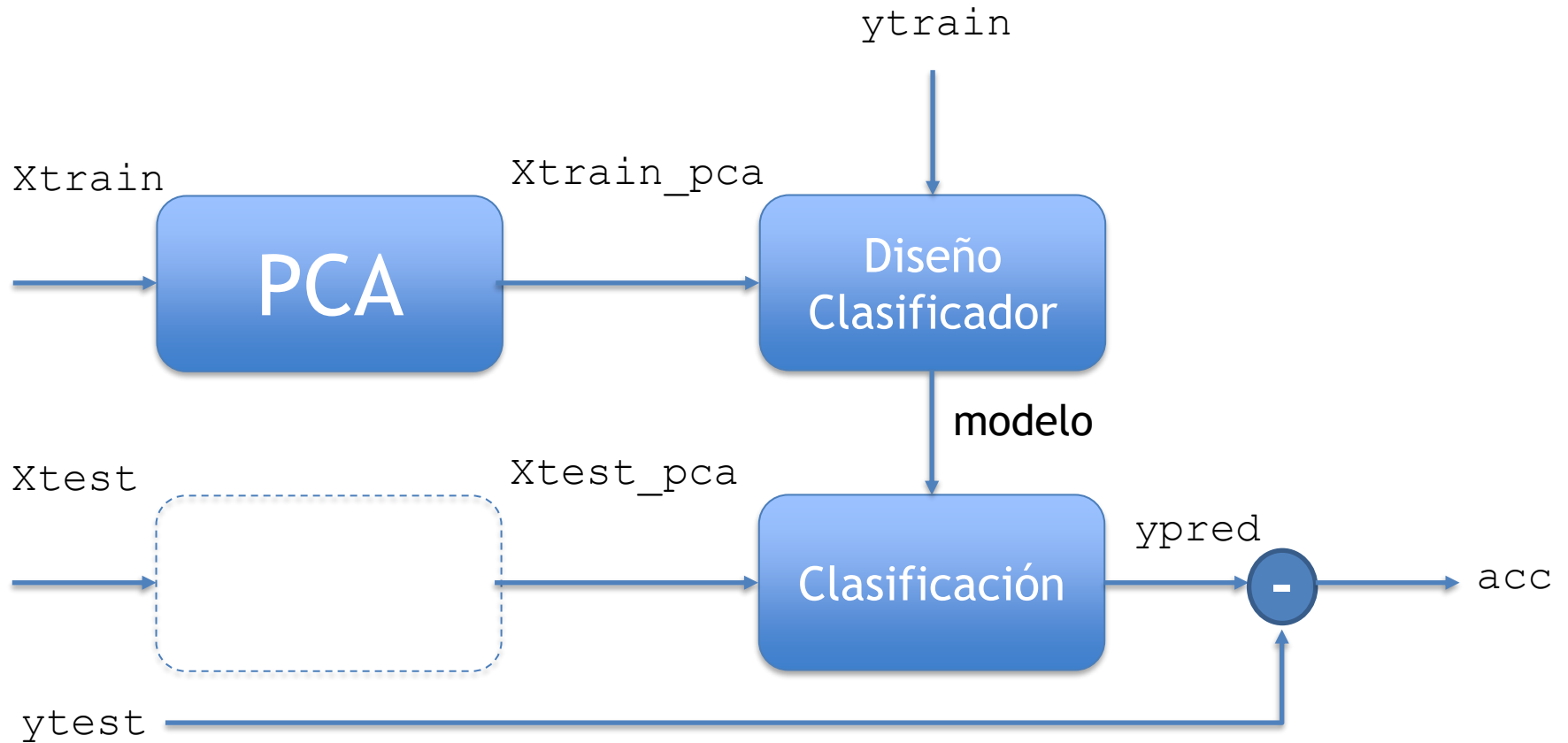
Características extraídas $N \times L$
(N muestras y L características)

Nuevas características $N \times L$
(N muestras y L nuevas características)
Se escogen las p primeras (las componentes principales)

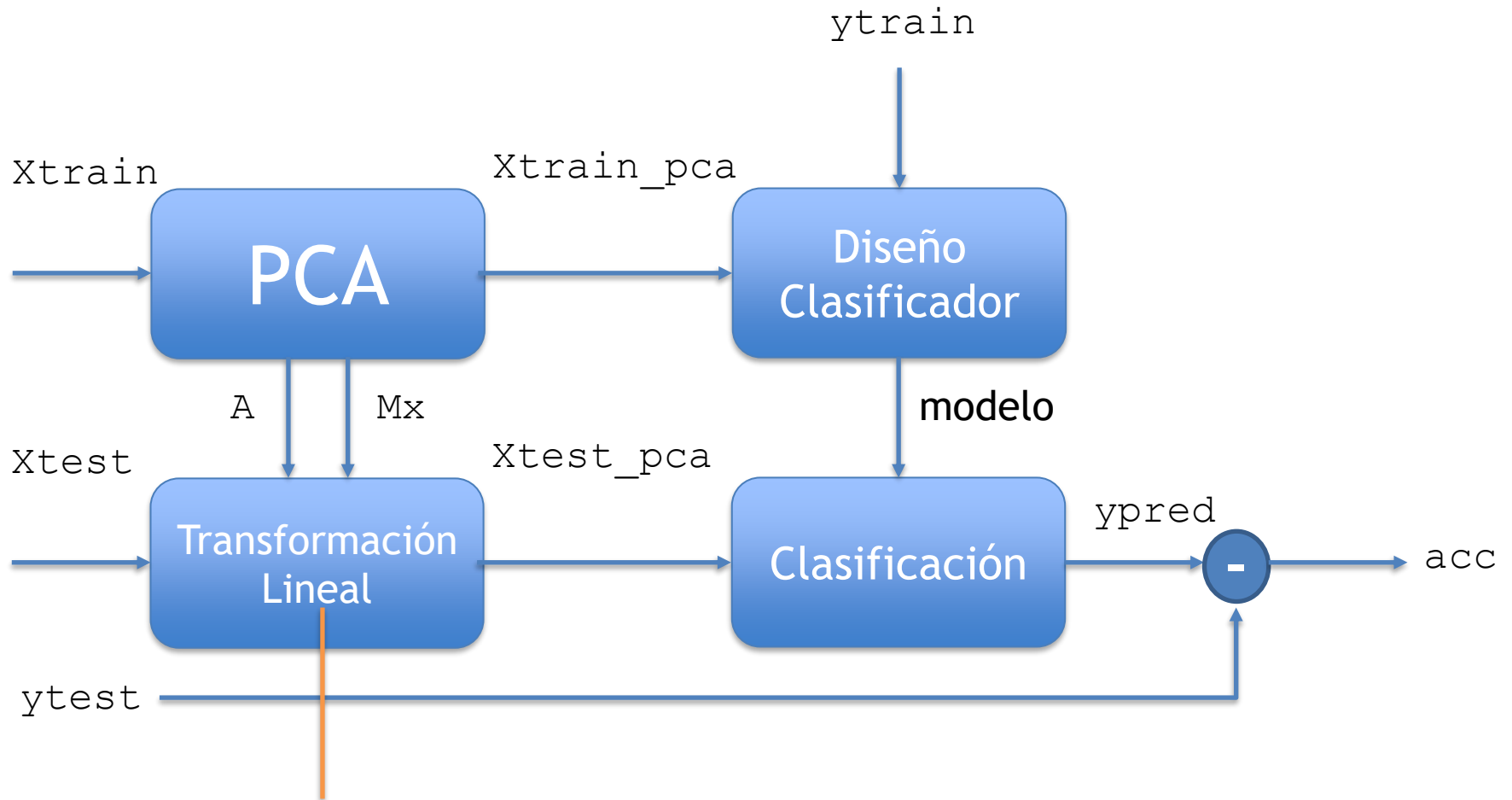
Training



Training / Testing



Training / Testing



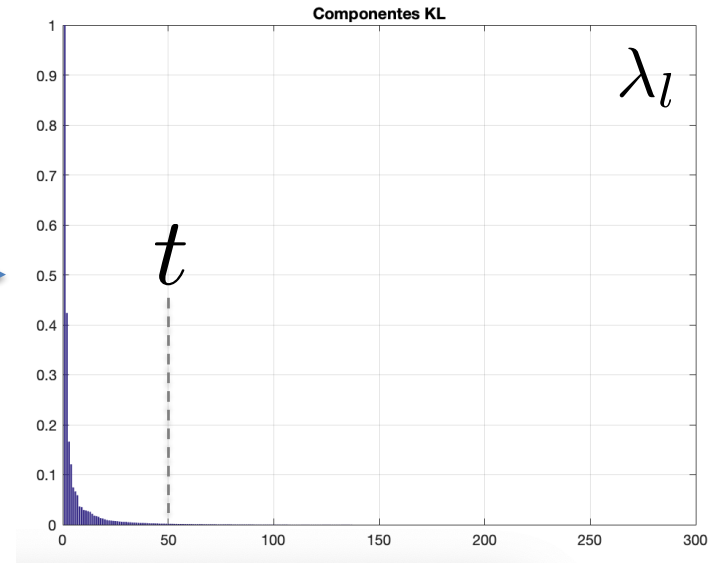
Ventajas de PCA:

1. La representación de muchas características es mucho más sencilla ya que se reduce la dimensionalidad
2. La clasificación usando las primeras componentes es más simple y más rápida para el clasificador
3. Es posible reconstruir \mathbf{X} a partir de las principales columnas de \mathbf{Y} . $\mathbf{X} = \mathbf{Y}\mathbf{A}^T + \mathbf{M}_x$

Reconstrucción con PCA:



PCA



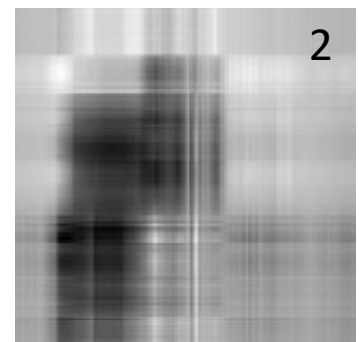
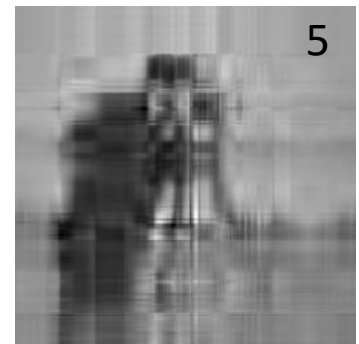
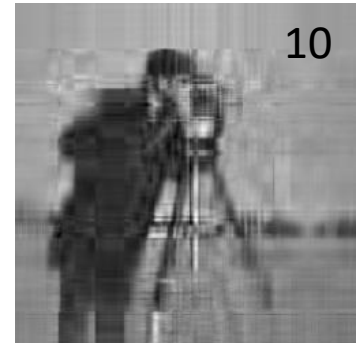
$\mathbf{A}_t \leftarrow$ primeras t columnas de \mathbf{A}

$\mathbf{Y} = \mathbf{X}_0 \mathbf{A}_t \leftarrow$ primeras t componentes principales de \mathbf{X}

$\hat{\mathbf{X}} = \mathbf{Y} \mathbf{A}_t^T + \mathbf{M}_x \leftarrow$ reconstrucción de \mathbf{X}

Reconstrucción con PCA:

t	Error
256	0.00%
200	0.08
100	0.94%
50	2.05%
30	2.93%
20	3.67%
10	5.36%
5	6.98%
2	9.10%

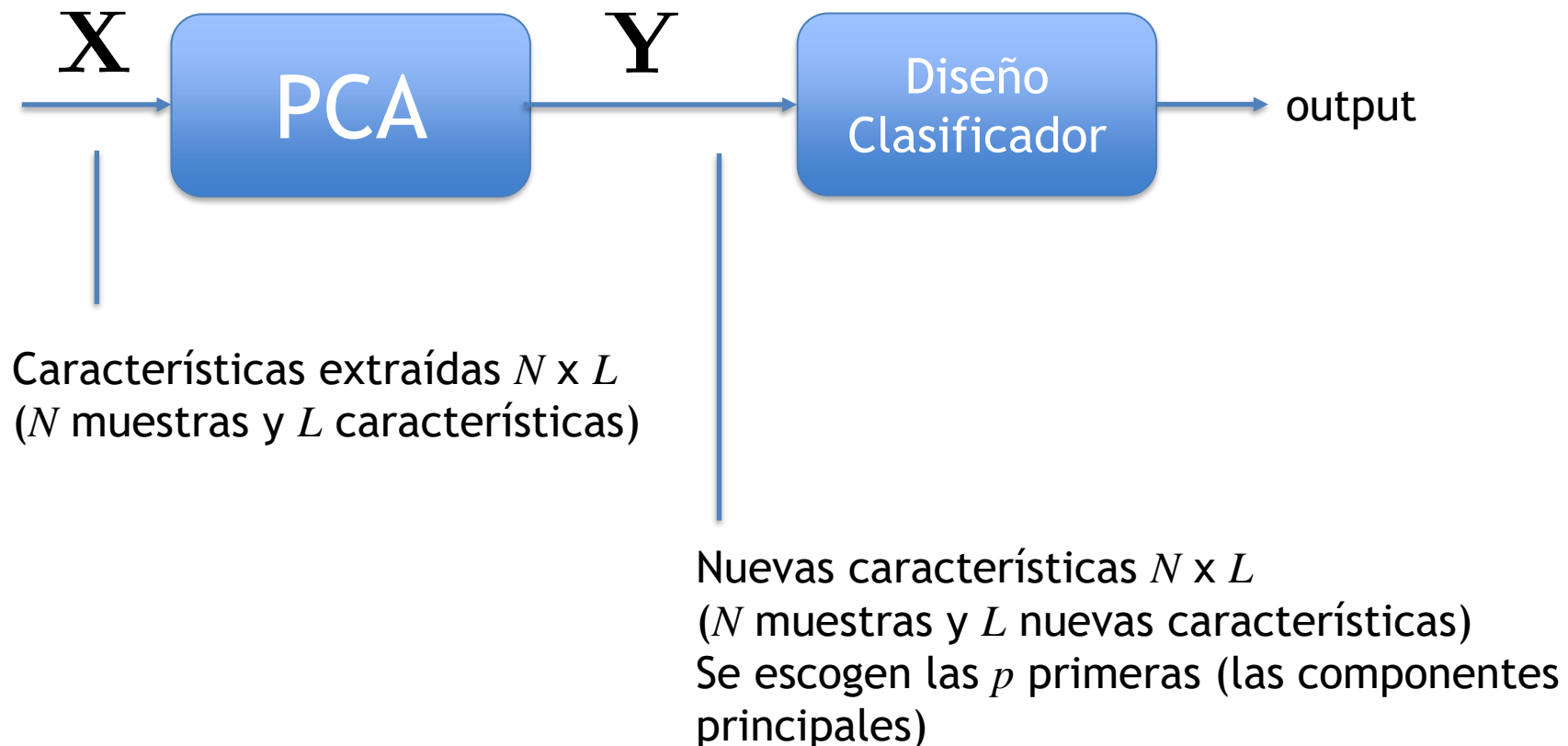


Desventajas de PCA:

1. Las características de PCA dependen de todas las características extraídas (es necesario extraer todas las características X para obtener Y , no hay ahorro de cómputo en la extracción)
2. PCA al no incluir información de los *labels* no asegura que las características transformadas tengan una buena separabilidad.

ALTERNATIVAS

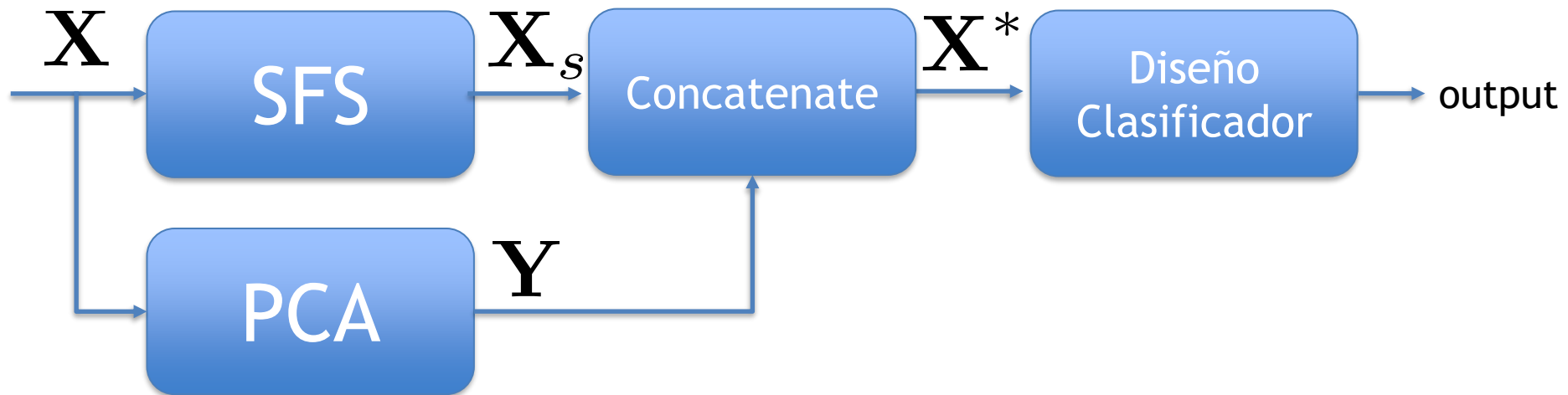
Estrategia con PCA (original):



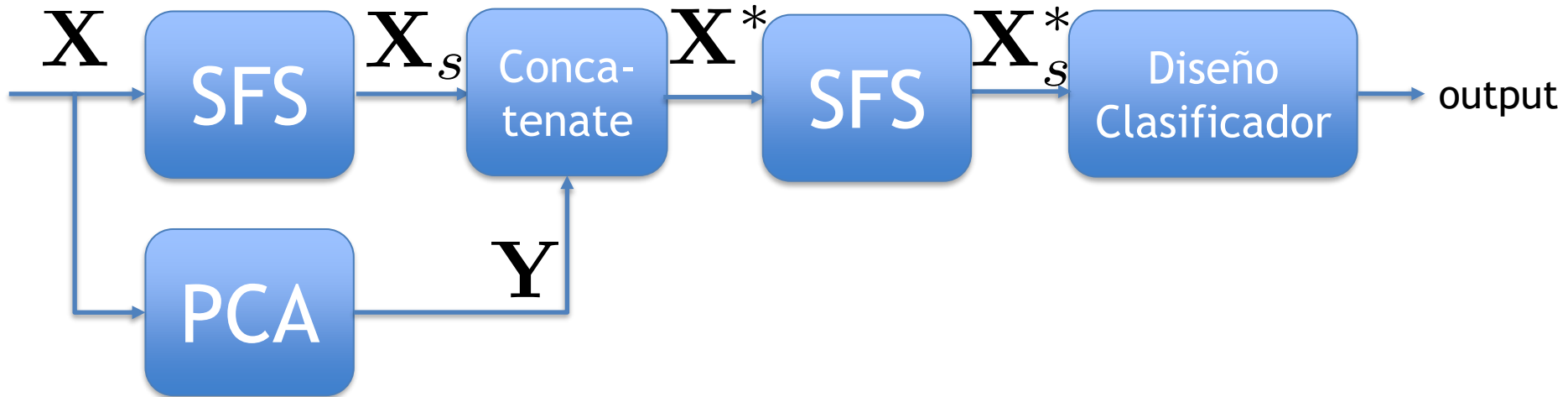
Estrategia con PCA y SFS (1 / 3):



Estrategia con PCA y SFS (2/3):



Estrategia con PCA y SFS (3/3):



OTRAS TRANSFORMACIONES

ICA

(Independent Component Analysis)

Diferencia de ICA con PCA:

- PCA elimina correlaciones pero no otro tipo de dependencia.
- ICA elimina correlaciones y dependencias de mayor orden.
- En PCA se persigue que la covarianza entre las variables transformadas sea 0

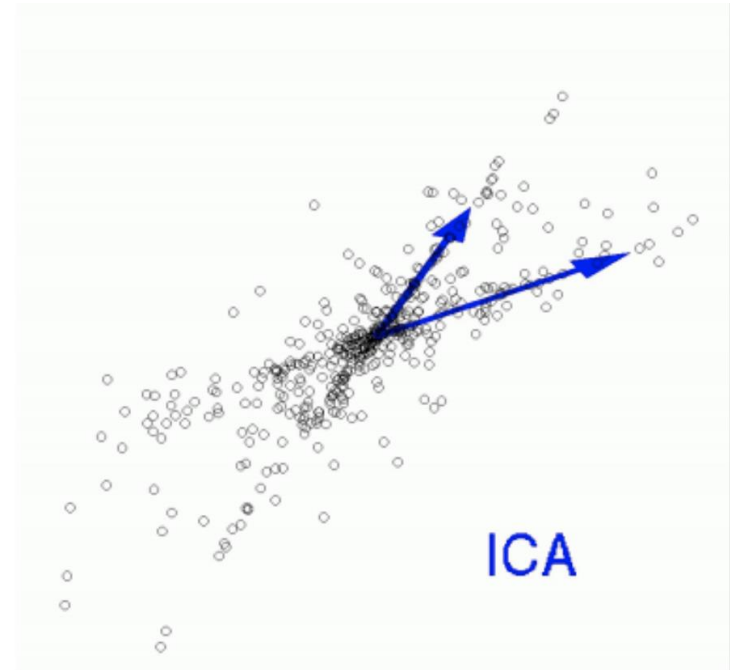
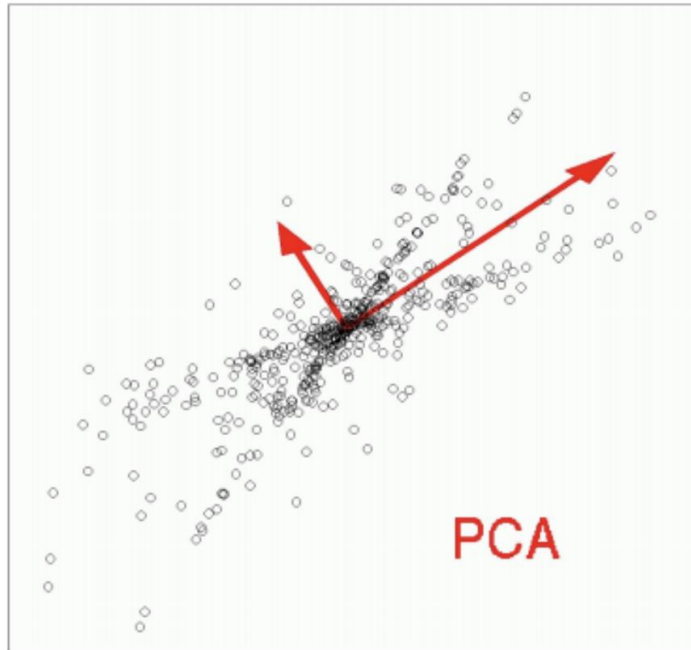
$$\text{cov}(y_k, y_l) = 0$$

- En ICA se persigue que las variables sean independientes:

$$p(y_1, y_2, \dots, y_L) = p(y_1)p(y_2) \cdots p(y_L)$$

Diferencia de ICA con PCA:

- En PCA los vectores son ortogonales.
- En ICA los vectores no son ortogonales.



PLSR

(Partial Least Squares Regression)

Descomposición PLSR:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \quad \mathbf{X}: \text{Matriz de características}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F} \quad \mathbf{Y}: \text{Matriz de labels}^{(*)}$$

Las descomposiciones de \mathbf{X} e \mathbf{Y} se hacen con el fin de maximizar la covarianza entre \mathbf{T} y \mathbf{U} .

Las matrices \mathbf{P} y \mathbf{Q} son ortogonales. Las matrices \mathbf{E} y \mathbf{F} corresponden al error.

La idea es buscar \mathbf{T} , que es la transformada de \mathbf{X} .

^(*) Matriz binaria de $N \times K$, $Y(i,k) = 1$ if $y(i) = k$, con K número de clases